

of a letter. Clearly, much of this type of 'staging' is non-linguistic (part of the external context of discourse), but its effect on our interpretation of the text of letters and many other types of discourse should not be ignored.

5 Information structure

5.1 **The structure of information**

In the previous chapters we have been considering increasingly restricted views of the production and interpretation of discourse. In Chapter 2 we considered the effect of situational context on discourse and in Chapter 3 the effect of different perspectives of topic structure. We devoted Chapter 4 to discussing the effect of linearisation in discourse, how what is presented first limits the interpretation of what follows and how decisions on thematisation provide the overall structure within which the addressee interprets the discourse.

In this chapter we focus in even further, to the smallest units of discourse structure: small local units at the level of phrase or clause. We consider how information is packaged within such small structures and, particularly, what resources are available to speakers and writers for indicating to their addressees the status of information which is introduced into the discourse.

5.1.1 *Information structure and the notion 'given / new' in intonation*

The serious study of information structure within texts was instituted by scholars of the Prague School before the Second World War. They studied what they called 'the communicative dynamism' of the elements contributing to a sentence, within the framework of 'functional sentence perspective'. (For an overview of this work see Vachek, 1966; Firbas, 1974.)

Many of the insights developed by the Prague scholars were first brought to the attention of Western scholars by Halliday in an extremely influential article published in 1967. Halliday elaborated and developed those aspects of Prague work which related directly to his own interests in the structure of texts. In particular, he

adopted the Prague School view of information as consisting of two categories: **new information**, which is information that the addressor believes is not known to the addressee, and **given information** which the addressor believes is known to the addressee (either because it is physically present in the context or because it has already been mentioned in the discourse).

Halliday further followed the Prague School in supposing that one of the functions of *intonation* in English is to mark off which information the speaker is treating as new and which information the speaker is treating as given. In his discussion of information structure, Halliday is particularly concerned to specify the organisation of information within spoken English and to relate this organisation to phonological realisation, especially to intonation. More recently, many scholars have extended the discussion of 'given' and 'new' information to the range of syntactic structures which are held to realise these categories of information. This has resulted in a drift of the meaning of the terms so that their extensions, particularly that of 'given', are now very much wider than Halliday intended them to be and, more importantly, no longer relate to the intonational phenomena which Halliday was concerned to describe. Once again we have a potentially confusing range of meaning associated with one term. We shall attempt to keep the meanings distinct by organising our discussion in the following way. First, we shall outline Halliday's account of information structure and its intonational realisation. (We shall cite, almost exclusively, Halliday's 1967 paper since this is the paper which has most influenced the work of other scholars. Halliday's own position on some of the points we discuss has moved considerably since 1967, cf. Halliday, 1978.) We shall follow this outline with a critique of Halliday's position, and the statement of a somewhat different position, still on information structure as realised in intonation. After that, we shall turn our attention to the syntactic realisation of information structure and, at this point, we shall confront the change in meaning of the terms 'given' and 'new'.

5.1.2 *Halliday's account of information structure: information units*

Halliday assumes that the speaker intends to encode the content of the clause (the basic unit in his grammatical system). In

many ways what Halliday views as the 'ideational' content of a clause may be compared with what others have called the 'propositional' content of a simple sentence (see discussion in 3.7). This clause content is organised by the speaker into a syntactic clausal structure, in which the speaker chooses among the thematic options available to him and, in spoken language, the clause content is organised into one or more **information units** which are realised phonologically by intonation.

According to Halliday, the speaker is obliged to chunk his speech into information units. He has to present his message in a series of packages. He is, however, free to decide how he wishes to package the information. He is 'free to decide where each information unit begins and ends, and how it is organised internally' (1967: 200). Thus, given that the speaker has decided to tell his hearer that 'John has gone into the garden with Mary', the speaker may package this information into one chunk as in

(1) a. John has gone into the garden with Mary

or two or three chunks as in

b. John – has gone into the garden with Mary

c. John – has gone into the garden with – Mary

The realisation of this difference in chunking will be discussed in the next section.

The 'internal organisation' of the information unit, relates to the way in which given and new information is distributed within the unit. Characteristically, Halliday suggests, the speaker will order given information before new information. The 'unmarked' sequencing of information structure is taken to be *given-new*. Naturally, information units which are initial in a discourse will contain only new information.

5.1.3 *Halliday's account of information structure: tone groups and tonics*

Information units are directly realised in speech as **tone groups**. (Other descriptions have called units of the same general size 'breath groups', 'phonemic clauses', or 'tone units', cf. Lehiste, 1970.) The speaker distributes the quanta of information he wishes to express into these phonologically defined units.

Tone groups are distinguished phonologically by containing one, and only one, **tonic syllable**. The tonic syllable is characterised as having the maximal unit of pitch on it. (In other descriptions it is called 'nuclear syllable' or 'sentence stress' and is characterised as having maximal moving pitch, maximal pitch height, maximal intensity and / or maximal duration, cf. Lehiste, 1970.) Tone groups, being produced in spoken language, are also related to the rhythm of spoken language (as described in Abercrombie, 1964). In Halliday's terms, each foot begins with a stressed syllable and contains any number of following unstressed syllables. It follows that tone groups must begin with a stressed syllable. Occasionally, the first syllable in the initial foot in a tone group is unstressed. A **silent ictus** (equivalent to a silent 'beat' in music) is then postulated as initial in the tone group. In the following example the tonic is marked by capitalisation, tone group boundaries by //, and the silent ictus by ^:

// ^ I / find it incompre / HENsible //

The tonic syllable functions to focus the new information in the tone group. In the unmarked case, the tonic syllable will focus the last lexical item in the tone group, which will generally be the head-word of the constituent containing new information. Consider the recital by a four-year-old of a fairy story which is very well known to her:

- (2) a. // ^ in a / far-away / LAND //
 b. // ^ there / lived a / bad / naughty / FAIRy //
 c. // ^ and a / handsome / PRINCE //
 d. // ^ and a / lovely / PRINcess //
 e. // ^ she was a / really / WICKed / fairy //

The child (no doubt influenced by the read-aloud version which she has heard, which in turn will be influenced by the punctuation of the written version) chunks her story into information units which are realised as tone groups. In tone groups *a-d*, the last lexical item receives a tonic syllable, which marks this as the focus of new information. In tone group *e*, the tonic syllable does not fall on the last lexical item, *fairy*, since the 'fairy' is already given in the preceding co-text and is treated as given by the speaker. The tonic syllable falls on the last lexical item which indicates 'new' information, on *WICKed*.

It is important not to suppose that the status of information is dictated by whether or not an entity has been referred to already within the discourse. As Halliday consistently and correctly remarks: 'These are options on the part of the speaker, not determined by the textual or situational environment; what is new is in the last resort what the speaker chooses to present as new, and predictions from the discourse have only a high probability of being fulfilled' (1967: 211).

Halliday states that there is a close relationship between the realisation of the information unit phonologically, in the tone group, and syntactically, in the clause: 'In the unmarked case (in informal conversation) the information unit will be mapped on to the clause, but the speaker has the option of making it co-incide with any constituent specified in the sentence structure' (1967: 242). (3) is an extract from a transcription presented by Halliday (1967: 201), where the speaker organises his information into phonological chunks, tone groups, which are co-extensive with both clauses and phrases:

- (3) // I had one of those nice old tropical houses // I was very lucky // it was about thirty years old // on stone pillars // with a long staircase up // and folding doors // back on to a verandah //

Compare (2) where the child also organised her story into tone groups which are co-extensive with clause and phrase.

In the next two sections we shall return to a consideration of the following aspects of Halliday's model:

- (a) the nature of the category 'tone group'
 (b) the nature of the category 'tonic'
 (c) the relationship of the information unit to phonological and syntactic categories.

5.1.4 Identifying the tone group

It is clear in listening to much unplanned spontaneous speech that speakers produce units which are rhythmically bound together, which are not always readily relatable to syntactic constituents, but which appear to be intended by the speaker to be taken together. It seems reasonable to call these, as Halliday does, 'information units'. If the discourse analyst wishes to characterise the realisation of information units, he needs an analytic system

which enables him to recognise these realisations in a reliable and principled manner. In working with speech read aloud, or with previously rehearsed speech, it is often possible to identify tone groups in the stream of speech, particularly as syntactic boundaries regularly coincide with phonological boundaries. However, in unplanned spontaneous speech, there are problems in identifying the tone group by phonological criteria alone. In principle, if tone groups really can be distinguished by phonological criteria alone, it should be possible to identify them from a content-indecipherable, but tone-clear, recording. In practice it is not. The claim, as expressed, is too strong.

Halliday describes the intonational contour of tone groups as being constituted around the tonic syllable: 'Within the tone group there is always some part that is especially prominent . . . The tonic syllable carries the main burden of the pitch movement in the tone group' (1970a: 4). The clear indication is that there is just one strong intonational movement within the tone group. It is possible to find such smoothly articulated intonational contours, but they are comparatively rare. It is usual to find tightly rhythmically bound structures with several peaks of prominence. Brown, Currie & Kenworthy (1980) report a series of experiments in which judges, experienced in teaching Halliday's system, were unable to make reliable identifications of tonics, hence unable to identify tone groups reliably.

If it is frequently difficult or impossible to identify the single peak of prominence round which a tone group is structured, it ought to be possible, in principle, to locate the boundaries of the unit. Halliday does not specify criteria for identifying the boundaries. He does, however, indicate that the boundaries will be, in part at least, determined by the rhythmic structure of the utterance: 'the tone group is a phonological unit that functions as realisation of information structure. It is not co-extensive with the sentence or the clause or any other unit of sentence structure; but it is co-extensive, *within limits determined by the rhythm*, with the information unit' (1967: 203, our emphasis). This insistence on tying the information unit directly to the form of phonological realisation yields some odd-looking information units as in (4).

- (4) // not only THAT but you // didn't know / where to start /
LOOKing for the / other and a // GAIN as I / say . . .
(1967: 209)

A similar information unit boundary located in the middle of a word occurs in another example transcribed by Halliday, cited as (5) below. These tone group boundaries seem counterintuitive if they are really to be regarded as the direct encoding of the boundaries of information units in speech. There are problems, then, with the identification of tonics and of tone groups in spontaneous speech. We offer an alternative system of analysis in 5.1.5 and 5.1.7 which, naturally, confronts the same sort of difficulties, but does, we believe, offer the practical discourse analyst a more secure basis for the identification of its categories.

5.1.5 The tone group and the clause

A problem arises from Halliday's commitment to the clause as the principle unit of syntactic organisation. In spite of his assertion that 'in the unmarked case (in informal conversation) the information will be mapped on to the clause', in his own extended transcription of conversation, the *phrase* seems a much more likely candidate:

- (5) // ^ it was a / fast / line and // very pleasant to travel on // much
the most / interesting / route to the / north // ^ but for / some /
reason it was ne // glected . . . //

(1970a: 127)

Extracts (2) and (3) (also quoted from Halliday) exemplify the same phenomenon. The point for concern here is whether in a model which does not take the clause as the domain of the main areas of syntactic choice, as Halliday's does, it is helpful or necessary to identify units which are mapped on to phrases rather than clauses as 'marked'. If they are 'marked', and the marking means anything, then some special, implicated, meaning ought to attach to them. If we simply look at paradigm sentences cited out of context as in (1a and b), it seems reasonable to suggest that

- (1b) // John // has gone into the garden with Mary //
is in some sense more 'marked' (nudge, nudge) than
(1a) // John has gone into the garden with Mary //

If, however, instead of using a sentence which is open to 'significant' interpretations quite independently of intonation, we use a less weighted sentence, say (6),

(6) // the room // is taking a long time to warm up //

the 'marked' significance seems to disappear. It is possible that intonation, together with pausing and other paralinguistic features of voice quality, may contribute to a 'marked' interpretation. It has yet to be shown that presenting phrases, rather than clauses, as information units does contribute to a marked reading. In the absence of such a demonstration, we shall not recognise an unmarked syntactic domain for information units.

If we abandon the clause as the unmarked syntactic domain of the information unit, more follows. You will remember that Halliday suggests that the unmarked structure of information within the information unit will be that *given* information precedes *new* information. This is very plausible if you take the clause (or simple sentence) as the unmarked syntactic domain because you can choose your examples from genres like those illustrated in (11) to (16) in Chapter 4 (obituaries, encyclopaedic entries, etc.) where, indeed, you are likely to find a 'given' form, referring to the topic entity, at the beginning of a clause, which is then followed by new information. Indeed you can sometimes perceive this organisation in snippets of conversation as in (7).

(7) we didn't see snow till we came up - the motorway

where 'we' is given in the context of discourse. If, however, we look at *phrases* marked out as information units, it is rarely going to be the case that they contain given information, unless the whole phrase is given. (Consider the phrases marked as information units in (2) and (3).) In information units realised as phrases, then, we are more likely to find all new information.

We return to the question of the syntactic unit which realises information structure in 5.2.1.

5.1.6 *Pause-defined units*

A number of people working on intonation in discourse have found a problem with the principled identification of tone groups by phonological criteria alone and have resorted to working

with units bounded by pauses in the stream of speech (for example, Chafe, 1979; Brown, Currie & Kenworthy, 1980; Butterworth (ed.), 1980; Deese, 1980).

The use of pause phenomena as a basis for building an analysis of chunking in spoken discourse might, at first glance, seem a rather precarious undertaking. The number and duration of pauses used by a speaker will obviously vary according to his rate of speech. It would be unlikely then, that one particular pause length, say one second, would have a single function for all speakers in all speech situations (a problem familiar to anyone who has worked with any kind of phonological data). However, one obvious advantage of working with pauses is that they are readily identifiable and, apart from the very briefest 'planning' pauses, judges have no difficulty in agreeing on their location. They are, furthermore, amenable to instrumental investigation, hence, measurable. What we might hope to find, in investigating the incidence of pauses, is different types of pauses in some regular pattern of distribution.

In an investigation of this type, it is important to choose the data used in the investigation so that you can generalise across speakers. A further practical point is that you should work initially on data which, you believe, will yield regular units which you will be able to recognise. If you begin by working in this tricky area with uncontrolled 'once-off' data you may find yourself working with data like that exemplified in (8) in which it is hard to discern regularities.

- (8) (numbers indicate pause lengths in seconds)
- a. but (0.8)
 - b. as (0.3) is well known (1.1)
 - c. it (0.2) very frequently happens that you you'll get a (0.3) co-occurrence of (0.2) an item with (1.0)
 - d. a recognised grammatical class (1.0)
 - e. erm (0.4) say a class of possessives or a class of (0.7)
 - f. erm negatives of one sort or another (0.6)
 - g. so that it is (0.4) erm (1.1)
 - h. we cannot restrict lexical patterning (0.6)
 - i. entirely to items as items (1.1)
 - j. erm (0.6)
 - k. whatever that may mean in itself may mean (1.2)

This extract from a post-graduate seminar presents a sample of speech very near the beginning of the seminar before the speaker

has 'got into his stride'. It is a perfectly familiar phenomenon and one which produces particularly 'dis-fluent' speech. (Even here certain regularities might tentatively be identified. The very brief pauses (0.2-0.3) are barely perceptible. There are four occurrences of pauses immediately adjacent to *erm*, a conventional 'planning marker'. Pauses regularly occur following 'sentence' boundaries (*d* and / or *e*, *h* and *k*.) The analyst makes life considerably easier for himself if he works, initially, with speech where he knows what it is the speaker is trying to say, and where he can make direct comparison across speakers (a methodology exploited for a range of purposes in, for example, Linde and Labov, 1975; Grosz, 1981; Chafe (ed.), 1980; Levelt, 1981).

In a study of the speech produced by twelve pairs of undergraduates, in which one member of the pair described a diagram which he could see, but his listener could not, so that his listener could draw the diagram, we were able to observe the incidence of pauses in comparable speech across a number of speakers. A typical piece of speech produced under these conditions is shown in (9).

- (9) A: halfway down the page (0.3) draw (0.6) a red (0.4) horizontal line (0.2) of about (0.5) two inches (16) on eh (1.1) the right hand side just above the line (1.9) in black (0.1) write ON (3.2)
 B: ON (3.4)
 A: above the line (1.4) draw (0.2) a black (0.65) triangle (1.0) eh (1.9) a right-angle (0.2) triangle (1.9) starting to the left (0.2) of the red line (1.0) about (0.9) half a centimetre above it (4.0)

In extract (9) the following pause types, defined in terms of relative length, can be identified.

1. *Extended pauses* These are long pauses which, in this extract, extend from between 3.2 to 16 seconds (which occur at points where the speaker has provided sufficient information for the hearer to draw or write what has been described). We represent such pauses in our transcriptions by ++.
2. *Long pauses* These pauses range from 1.0-1.9 seconds in this extract. We represent such pauses by +.
3. *Short pauses* These pauses range between 0.1-0.6 seconds in this extract. We represent such pauses by -.

We assume that extended and long pauses might be proposed as unit boundaries, whereas the short pauses might be treated as unit-internal. Adopting this view (9) can be presented as (10):

- (10) A. halfway down the page - draw - a red - horizontal line - of about - two inches ++
 on eh + the right hand side just above the line + in black - write ON ++
 B. ON ++
 A. above the line ++
 draw - a black - triangle + eh + a right-angle - triangle + starting to the left - of the red line + about + half a centimetre above it ++

(In 5.1.8 we shall consider the structure of information in data of this sort and in 5.1.7 we shall consider the role of phonological prominence in the same set of data.)

The ranges of pause lengths found across subjects in this data is summarised in (11).

(11)

0	0.5	1	1.5	2	2.5	3
└──────────┘		└──────────┘		└──────────┘		
short		long		extended		
pauses		pauses		pauses		

We represent pauses as following utterances, as though pauses represented termination markers in the way punctuation represents termination markers. We should note that Chafe (1979), whose work on pauses shows results very similar to ours, indicates pauses as preceding utterances, since he regards the pause length as a function of the amount of planning which the speaker is putting into his next utterance. Chafe reports work in which subjects are asked to retell a series of events which they have seen enacted in a short silent film. Chafe observes 'major hesitations' (equivalent to our 'extended pauses')

... at those places in a narrative where paragraph boundaries seem to belong. For example, look at the following portion of one of our film narratives, in which the speaker finishes telling about a boy's theft of some pears and begins telling about some other things that happen subsequently (Speaker 22):

And as he's holding onto the handlebars he takes off with them
(1.1) Um - (0.7) then (0.4) uh - (2.1) a girl - on a bicycle
(1.15) comes riding towards him . . . in the opposite direction.

. . . *Um-* and *uh-* are lengthened pronunciations of . . . "pause fillers". The total amount of time spent hesitating between the end of the first sentence and the beginning of *a girl on a bicycle* was 6.25 seconds. We have here clear evidence of some important and time-consuming mental processing. (1979: 162)

This view of pauses as indicating the time spent by the speaker on constructing the following utterance is obviously interesting. In our discussion of how the discourse analyst may proceed in analysing discourse we shall take the more sober view of extended pauses and long pauses as constituting boundaries of phonological units which may be related to information units.

5.1.7 *The function of pitch prominence*

Halliday makes the simplifying assumption that there is only one function of pitch prominence, 'the main burden of the pitch movement', and that is to mark the focus of new information within the tone group. In fact the limited resources of pitch prominence may mark a good deal more than this. They are also exploited by speakers to mark the beginning of a speaker's turn, the beginning of a new topic, special emphasis, and contrast, as well as information which the speaker presents as new. In our view, phonological prominence (which may vary considerably in phonetic realisation from one accent to the next) has a general *watch this!* function and, *inter alia*, is used by speakers to mark new information as requiring to be paid attention to. Phonological non-prominence is then associated with all the elements which the speaker does not require the hearer to pay attention to, which will include not only 'given' information, but also, for example, unstressed grammatical words. (For an extensive discussion of this point of view, cf. Brown, Currie & Kenworthy, 1980; Yule, 1980.)

We have already remarked (5.1.4) that experienced judges, in a series of experiments, were not able to identify 'tonics' consistently. Whereas some judges did show a consistent preference for the last lexical item in a phrase (which tends to be realised with extended length), any lexical item which introduced new information in the discourse was recognised by some of the judges, some of the time,

as 'a tonic'. In an extensive series of instrumental measurements, it was possible to show that the phonetic cues which have traditionally been claimed to mark the tonic (maximal pitch movement, maximal pitch height and maximal intensity) rarely cumulated on one word in spontaneous speech (except in cases of contrast), but tended to be distributed separately or paired, over words introducing new information. Where these maxima competed as cues, judges regularly identified several tonics in short phrases, even two tonics in two-word phrases.

Such a finding is not, of course, inconsistent with Halliday's remark that 'the speaker has the option' of relating the information unit to 'any constituent specified in the sentence structure'. What it does yield, however, is a density of identified tonics, hence tone groups, which is not indicated in any of Halliday's transcriptions of speech, and it certainly destroys the notion of the unmarked relationship between clause (or indeed phrase) and tone group. What it leaves is a notion that any lexical item introducing new information (new entity-noun, new property-adjective, new activity-verb) is likely to be realised with phonological prominence, which is likely to be identified as a 'tonic'.

Many scholars working on intonation, particularly those working on intonation in conversational speech, have abandoned (if they ever held) the requirement that information units, however realised, should contain only one focus, hence be realised with only one tonic (cf., for example, Bolinger, 1970; Crystal, 1975; Chafe, 1979; Pellowe & Jones, 1979; Thompson, 1980). We, too, abandon the notion of a single tonic realising the focus of an information unit. We can now proceed to a more satisfactory representation of our four-year-old's fairy story (12).

- (12)
- a. in a FAR-away LAND +
 - b. there LIVED a BAD NAUGHTy FAIRy ++
 - c. and a HANDsome PRINCE +
 - d. and a LOVELy PRINCESS ++
 - e. and she was a REALly WICKed fairy ++

We have capitalised the phonologically prominent syllables in this representation. Hereafter we shall represent the phonologically prominent *word* as prominent, disregarding what is, for our present purposes, the irrelevance of the phonological structure of the word. (In this rendering, spoken by a child with a Yorkshire accent, the

final word in each information unit, whether or not it is phonologically prominent, is realised on an extended falling tone. Each preceding phonologically prominent word is realised on a less extended falling tone. Had the speaker been a Glasgow speaker, most, if not all, of the falls would have been rises.)

In the study of diagram descriptions, which we have already mentioned (5.1.6), we examined the distribution of phonological prominence with respect to information which we knew was being introduced into the discourse for the first time and with respect to information which we knew had already been introduced. Expressions introducing new information were realised with phonological prominence in 87% of cases as in:

- (13) a. draw a BLACK TRIANGLE
 b. draw a STRAIGHT LINE
 c. write OUT in BLACK
 d. there's a CIRCLE in the MIDDLE

Expressions introducing given information were produced without phonological prominence in 98% of cases (the exceptions were contrasts), as in:

- (14) (expressions mentioning given information are italicised)
 a. UNDERNEATH *the triangle*
 b. at the END . . . of *this line* write the word ON just ABOVE *the line*
 c. a LINE . . . about TWO INCHES + and ABOVE *it* write ON

How are we to account for the fact that the mention of some entities which we know independently are being introduced into the discourse for the first time are not being marked by the speaker with phonological prominence? Most of these examples apparently arise from the same source – although the entity introduced is certainly a new entity, the form of expression used to introduce it has just been used in the discourse to mention an immediately preceding entity. Thus one speaker mentions 'the triangle' and then introduces a different, new, triangle, speaking with a near monotone, low in her pitch range: *a right-angled triangle like the black one*. We could account for this lack of phonological marking on the grounds that the speaker believes that the hearer is, at the relevant point in the discourse, thoroughly acquainted with the task and

expects another triangle to be introduced, that triangles are 'in his consciousness'. Schmerling (1974: 72) suggests that, depending on very general contextual expectations, speakers may choose to mark different aspects of their message as new. Thus, she suggests that a speaker identifying a man in the street to his hearer may produce

This is the DOCTOR I was telling you about

but if the speaker is working in a hospital and talking to another employee he is more likely to produce

This is the doctor I was TELLING you about.

Similarly if you have already been describing one triangle and triangles are clearly 'in the air', you might not choose to mark yet another member of the same class as new. This seems at least a plausible explanation for the phenomenon we observe.

It does, however, lead on to a further and more general observation. This is that we should not expect 100% direct mapping from categories at one level of description into categories at another level of description. We have already observed the curious information units which Halliday is obliged to identify as a result of mapping them directly on to tone groups whose boundaries are determined by the rhythm of speech (see discussion in 5.1.4). It is our general experience in other areas of linguistic analysis that there has to be some accommodation between analytic units at one level of description and those at another. Thus the proposition identified at the level of semantic form is not directly mapped into syntactic units without syntactic residue. The English phonological unit / b / may be identified at a phonological level as a 'voiced bilabial stop', but at a phonetic level its realisation may sometimes be described as 'voiceless' or as 'fricative'. We should not expect things to be different at the level of information structure. It seems reasonable to suggest that information structure is realised partly by syntax (in the thematic structure, i.e. by word order) and partly by phonological systems including phonological prominence and pause. We should expect to find regularities in the realisation of information structure within these systems. We should not, however, be misled by the vast volume of work on the intonation of sentences read aloud, into supposing that there are categorial rules which map information units on to syntactic units which are

'Tis here we shall discover the use of the two Articles (A) and (THE). (A) respects our *primary* Perception, and denotes Individuals as *unknown*; (THE) respects our *secondary* Perception, and denotes Individuals as *known*. To explain by an example. I see an object pass by, which I never saw till then. What do I say? *There goes A Beggar, with A long Beard*. The Man departs, and returns a week after. What do I say then? *There goes THE Beggar with THE long Beard*. The Article only is changed, the rest remains unaltered.

(Harris, 1751: 215-16)

In (16) we exemplify a range of syntactic forms which have been frequently identified in the literature as expressions referring to given entities. The expression claimed to be 'given' is italicised in each case.

- (16)
- a. 1. Yesterday I saw a little girl get bitten by a dog.
2. I tried to catch *the dog*, but *it* ran away.
(Chafe, 1972: 52)
 - b. 1. Mary got some beer out of the car.
2. *The beer* was warm.
(Haviland & Clark, 1974: 514)
 - c. 1. Mary got some picnic supplies out of the car.
2. *The beer* was warm.
(Haviland & Clark, 1974: 515)
 - d. 1. Yesterday, Beth sold her Chevy.
2. Today, Glen bought *the car*.
(Carpenter & Just, 1977a: 236)
 - e. 1. I bought a painting last week.
2. I really like *paintings*.
(Chafe, 1976: 32)
 - f. 1. Robert found an old car.
2. *The steering wheel* had broken off.
(Clark, 1978: 310)
 - g. 1. What happened to the jewels?
2. *They* were stolen by a customer.
(van Dijk, 1977: 120)
 - h. 1. I saw two young people there.
2. *He* kissed *her*.
(Sgall, 1980: 238)
 - i. 1. (Sag produces a cleaver and prepares to hack off his left hand)
2. *He* never actually *does it*.
(Hankamer & Sag, 1976: 392)
 - j. 1. Look out.
2. *It's* falling.
(Carpenter & Just, 1977a: 236)

- k. 1. William works in Manchester.
- 2. So *do I*.

(Allerton, 1975: 219)

The syntactic forms which are regularly discussed in association with 'given' information include:

- A.
 - (i) Lexical units which are mentioned for the second time as in (16) *a* and *b*, particularly those in definite expressions.
 - (ii) Lexical units which are presented as being within the semantic field of a previously mentioned lexical unit as in *c*, *d*, *e* and *f*, again particularly those in definite expressions.
- B.
 - (i) Pronominals used anaphorically following a full lexical form in the preceding sentence as in *a*, *g* and *h*.
 - (ii) Pronominals used exophorically (to refer to the physical context of situation) where the referent is present, as in *i* and *j*.
 - (iii) Pro-verbals (less commonly discussed) as in *i* and *k*.

The discussions from which these examples are abstracted are concerned with particular syntactic realisations, in constructed sequences of sentences, where some element in the second sentence is held to be in some sense 'given'. We shall return in 5.3 to consider the range of interpretations of 'given' status which underlies these various sequences. For the moment we shall concentrate on the form of expressions which are considered as conventional indications that their referents are being treated by the speaker / writer as 'given'.

In the examples in (16) we find two predominant forms of expression used to refer to an entity treated as given, pronominals and definite NPs. These forms are often treated in the linguistics literature as though they were in free variation. It is clearly important for the discourse analyst to know whether the forms which occur in cited examples have, in fact, the same privilege of distribution in naturally occurring discourse as is attributed to them in invented sequences. If the forms are in free variation, one form should be as appropriate as another in the same context, so, for example, in (16g) the cited form:

(16) g. 2. *They* were stolen by a customer

might equally well be replaced by a repetition of the sort which occurs in (16b) yielding

g. 3. *The jewels* were stolen by a customer

Are these equally appropriate continuation forms? Are they understood in the same way?

It is clearly possible for the expressions used to 're-identify' a given entity to carry a good deal more information than that in (g3). Donnellan (1978) produces a sequence of forms illustrating this point, cited here as (17).

- (17)
- a. 1. A man came to the office today carrying a huge suitcase.
2. It contained an encyclopaedia.
 - b. 1. A man came to the office today carrying a huge suitcase.
2. The suitcase contained an encyclopaedia.
 - c. 1. A man came to the office today carrying a huge suitcase.
2. The huge suitcase carried by the man who came to the office today contained an encyclopaedia.

Donnellan comments 'In some of the cases repetition of information makes the discourse sound like the awkward language of a child's first reader' (1978: 58). If we find (17c) 'awkward', this is presumably because normally, in genres other than children's first readers, speakers do not reiterate so much 'given' information. A question which should interest discourse analysts is 'Under what conditions do these different forms occur?' Can we find instances in discourse of typically different distribution of pronominal forms as opposed to definite NPs?

We can give an answer to this question for one, limited, genre. It is important to be clear from the outset that the distribution we describe for this particular genre may not apply to all other genres. It is important to examine a wide range of data before beginning to make claims about 'typical distribution in English'.

The diagram-drawing data we have already referred to offers a clear advantage to the discourse analyst who is studying forms used to refer to new and previously mentioned entities, since he can distinguish, in his recording of the entire interaction, the occasions on which an entity is introduced for the first time, and the occasions on which an entity is subsequently mentioned. As each new entity

expression is introduced, it can be given a numerical subscript, as in (18):

(18) draw a *line*₁ and above *it*₁ write *ON*₂

As the analysis proceeds through the text, more new entities are introduced, and each succeeding expression introducing a new entity will receive a higher number. Thus in any particular stretch of analysed text, the NP with the highest number attached to it will mention the most recently introduced new entity. NPs with lower numbers attached to them, will be being used to refer to previously mentioned, given, entities as in the extracts in (19).

- (19)
- a. at the end of *the line*₂ draw a *square*₄
 - b. underneath *the red line*₄ write *IN*₇
 - c. draw a *diameter*₂ across *it*₁
 - d. draw a *straight line*₅ across *the circle*₁

Note that the expressions mentioning 'given' entities here include a pronominal expression *it*, a simple definite NP *the line* and a definite NP which specifies an identifying property *the red line*. Do all these forms occur in identical environments?

Armed with the method of indexing we have described, we can find out. We can distinguish an expression referring to the most recently introduced entity before the new entity. We shall call this the expression referring to a *current* given entity. Such expressions will always be indexed by the number immediately below that which marks the most recent 'new' expression. Thus in (19c) the expression *it* is used to refer to an entity which was introduced immediately before 'a diameter' was mentioned. Any expression indexed with a number greater than 1 below the number indexing the 'new' expression, we shall call an expression mentioning a *displaced* given entity. In (19 a, b and d) we have instances of 'displaced' expressions: *the line*, *the red line*, *the circle*. Consider now extract (20).

- (20) draw a *black triangle*₁ ++ underneath *the triangle*₁ draw a *red line* about two *inches*₂ ++ and at the right hand side of *this line*₂ write *ON* in *black*₃ ++

In the pause-defined unit *underneath the triangle draw a red line about two inches*, the expression *the triangle* mentions a current given entity, current with respect to the 'new' entity introduced by

the expression *a red line about two inches*. In the next unit, the expression *this line* mentions the current given entity, current with respect to the 'new' entity introduced by the expression *ON in black*.

With this distinction, current/displaced, established in the analysis, it is possible to check through the data and to examine the realisation forms in each case. In the restricted data we describe here, we find there is not free variation between all the forms we are interested in, as we show in (21).

(21)		<i>Current</i>	<i>Displaced</i>
	<i>the</i> + property + NOUN	3	44
	<i>the</i> + NOUN	21	56
	pronoun	65	
	ellipsis	11	

(In each case the figures represent the percentage of expressions under each head which have this particular form.)

It is clear that expressions of the form *the* + *property* + *noun* is used almost exclusively in identifying displaced entities. The simple definite NP, *the* + *noun*, is used predominantly in identifying displaced entities but is also used identifying current entities. In this data, *pronouns* never occur as expressions identifying displaced entities, only as expressions identifying current entities.

It must be clear that the norms of distribution we state for this limited data may be different in other genres. It should also be clear that the indexing analysis becomes a good deal more problematical when it is applied to richer data, as we shall briefly demonstrate in a discussion of a very different genre, the elliptical written language of a recipe.

(22)	<i>Chicken Korma</i>	
	<i>Ingredients</i>	
	1 roasting chicken, about	½ teaspoon mashed garlic
	2 lb, cut in pieces	5 oz plain yogurt
	½ lb butter or ghee	1-inch piece of ginger,
	1 medium-sized onion	sliced
	½ teaspoon ground saffron	1 teaspoon cloves, black
	3 teaspoons ground coriander	pepper and cummin seed,
	seed	whole
	½ teaspoon ground chilli	salt to taste

Method

Slice the onion finely, brown in the butter and then place in a small dish. Put the ground spices into a breakfast cup of water, add to the fat in the pan and cook for 3 minutes, stirring the while. Now add the chicken, mix well, see that the meat is just covered by water and boil for 20 minutes with the lid on the pan.

When the liquid has almost evaporated, continue to cook, but stir the chicken till golden brown.

Crush the browned onion with a spoon and add it to the chicken with the yogurt, salt to taste and remainder of the spices. Add a cup of water, put on the lid and simmer gently till the chicken is tender. (If the chicken is not quite done and the liquid has evaporated, add a little more water and cook for a further period.)

(Harvey Day, *The Complete Book of Curries*, The Cookery Book Club, 1970, p. 128)

Most 'new' entities are introduced into the discourse in the list of ingredients. The first mention under *Method* of items which were introduced in the list of ingredients, regularly takes the form of a definite lexical expression:

(23a) *the onion, the butter, the ground spices* (superordinate to those mentioned in the list), *the chicken, the yogurt*

We could characterise this writer's preference in discourse organisation as selecting an entity as 'topic entity' for a sequence of events within a sentence, and then producing no further mentions of that entity within the sentence, relying on ellipsis as in (23b).

(23b) Slice *the onion* finely, brown \emptyset in the butter and then place \emptyset in a small dish.

If we were to try to apply an 'index of mention' to the entities introduced in this sentence, the analysis would presumably appear as in (23c).

(23c) Slice *the onion*₁ finely, brown \emptyset ₁ in *the butter*₂ and then place \emptyset ₁ in a *small dish*₃.

This seems an unsatisfactory representation. Even if we were to accept the oddity of indexing ellipses, 'holes' in the sentence, the indexing does not appear to represent in a satisfactory manner what it is that is being referred to. The object which is 'placed in a small

dish' does not consist simply of 'the onion' but of 'the onion which has been browned in *the butter*'. As soon as data involving change-of-state predicates is studied, the simple indexing of mention which we applied in the diagram-drawing becomes inadequate. We need a much more complex mechanism capable of 'carrying' predicates and, indeed, of sloughing them gradually off, as the discourse proceeds. (We return to the general problem of 'pronominal substitution' and 'substitution by ellipsis' in 6.1.)

Whereas in the diagram-drawing data we had a secure basis for identifying first and subsequent mention which enabled us to draw a confident new/given and current/displaced distinction, this simple approach can no longer apply when we turn to richer data. Nonetheless the analysis developed for the restricted data has served its purpose since it enabled us to draw a principled distinction between 'current' and 'displaced' entity referring expressions. Generalising that distinction to (22) we can observe that displaced entities are regularly referred to by full lexical definite NPs.

In restricted data it is possible for the analyst to ascertain which expressions introduce new entities into the discourse and which expressions are used to introduce entities which have been specifically mentioned before. Once this has been done, the forms of those expressions become available for linguistic description. Observe that our interest does not lie simply in describing the form of the expression, which is obviously of prime interest to the sentence grammarian. Our interest lies in observing the forms *in the context in which they are used*. We want to know how speakers, having a given quantum of information to impart, identify and package that information.

5.2.2 Information structure and sentence structure

In the previous section, we examined the form of nominal expressions used to refer to entities on the occasion of their first and subsequent mentions. We pointed out that, as soon as the data becomes richer and involves change-of-state predicates, the secure distinction between first and second mention becomes blurred. It is no longer possible for the analyst simply to count through the mentions in a text and assert of an entity previously mentioned that it is, necessarily, 'given' (a point we also made with

respect to (15)). Halliday, as we saw in 5.1.5, has particularly insisted on the importance of taking this view: 'what is new is in the last resort what the speaker chooses to present as new, and predictions from the discourse have only a high probability of being fulfilled' (1967: 11).

In recent years, a number of psycholinguists working on information structure have taken the 'given/new' distinction as expounded by Halliday and applied it to written sentences, often written sentences cited in isolation. Since written sentences have no intonation, these writers assign intonation structure to them. They then rely on the syntactic form of nominal expressions, of the sort we were examining in the last section, and on sentence structure, to determine what, in the sentence, has the status 'new' and what has the status 'given'. They regard the information status as irrevocably assigned by the form of expression used. This approach to information structure has led to a re-interpretation of what is meant by the status 'given', as we shall see below and in 5.3.

Clark & Clark (1977: 93) report an experiment carried out by Hornby (1972). Hornby presented subjects with a series of written sentences read aloud. The sentences are reproduced in (24) as they appear in Clark & Clark's discussion:

(24) *GIVEN AND NEW INFORMATION*
Five types of sentences and their given and new information

SENTENCE	GIVEN AND NEW INFORMATION
1. It is the BOY who is petting the cat.	Given: X is petting the cat New: X = the boy
2. It is the CAT which the boy is petting.	Given: the boy is petting X New: X = the cat
3. The one who is petting the cat is the BOY.	Given: X is petting the cat New: X = the boy
4. What the boy is petting is the CAT.	Given: the boy is petting X New: X = the cat
5. The BOY is petting the cat.	Given: X is petting the cat New: X = the boy

The status of the capitalised elements is explained on page 32: 'Sentences signal given and new information by stress or accent on particular words (e.g. Halliday, 1967) . . . The word with the focal

stress, or a phrase containing it, always conveys the new information.'

A number of points arise from this presentation. First of all it is not entirely clear whether given / new status is being determined by the sentential form (as the sub-heading to the table suggests) or by the effect of the placing of 'focal stress' on different constituents of the sentences, or by some interaction of these two distinct systems.

Secondly, it somewhat misrepresents Halliday's position to suggest that he takes the view that '*Sentences* signal given and new information' (our emphasis). Halliday repeatedly insists that it is speakers who signal information status.

Thirdly, Hornby and, apparently, Clark & Clark are attributing one focus of information to each sentence. The only sentence in this set which is directly relatable to the Hallidaian clause is sentence 5. All the other sentences are realised in two clauses. In a Hallidaian analysis each of these two clauses would be expected to contain an information focus as is demonstrated in the similar examples discussed in Halliday (1967: 226):

- (25) a. // the one who painted the SHED last week // was JOHN //
 b. // JOHN // was the one who painted the SHED // last week //

By analogy we would expect sentence number 3, for example, to be realised with two points of information focus:

- c. // the one who is petting the CAT // is the BOY //

Fourthly the term 'given' is no longer being used as an analytic term to describe the status of the referents of expressions within the clause (or tone group), but is being used of the presuppositions attributed to clauses within sentences.

This view of 'givenness' is picked up by other psycholinguists – for instance, Sanford & Garrod, 1981. On page 92 they cite Halliday who, they say, 'goes on to suggest that the "Given-New" partition can apply to any sentence in a discourse and is signalled both in its syntax and intonation. To take a simple example, the sentence "It was Mary who left" seems to break down into Given: "Someone left"; and New: "The someone who left was Mary".' On page 93, they continue 'Now if we go back to our example, we see that this is naturally read as "It was MARY // who LEFT //",

meaning that MARY and LEFT are given stress and so treated as New information, whereas "It was" and "who" are not given stress and so treated as Given.' The analysis on page 92 resembles the Hornby / Clark & Clark analysis. The analysis on page 93 owes more to Halliday (though it is doubtful if he would consider *it was* and *who* as being 'treated as Given', cf. 1967: 229ff.). The first analysis is concerned with what is *presupposed* in the 'relative' clause and derives from processes of sentential thematisation, the second analysis is concerned only with the information status of the referents of expressions within the clause, and derives from imputed intonation structure. Both are perfectly reasonable types of analysis of different phenomena, whose categories may interact to produce a particular effect, but they must not be confounded.

In 5.3 we shall see how far this procedure of taking what is 'presupposed' in a sentence, and including it within the category of 'given information', has extended the meaning of the term 'given' in the psycholinguistics literature in such a way that it no longer clearly relates to the way the term is used in the linguistics literature.

5.3 The psychological status of 'givenness'

In this section, we turn from an examination of the forms, intonational and syntactic, which are used by speakers to indicate given / new status, to a consideration of what scholars using these terms have understood by them. How does information come to have the status of 'being treated as given'?

5.3.1 What does 'given' mean?

Halliday produced characterisations of given / new, in terms of speaker-expectations, which are capable of being rather narrowly interpreted, as we suspect he intended. 'Given' information is specified as being treated by the speaker as 'recoverable either anaphorically or situationally' (1967: 211) and 'new' information is said to be focal 'not in the sense that it cannot have been previously mentioned, although it is often the case that it has not been, but in the sense that the speaker presents it as not being recoverable from the preceding discourse' (1967: 204). Whereas this characterisation does discriminate between the status of information marked intonationally as 'given' or 'new' by the speaker,

it is capable of being interpreted so as to embrace a wider range of other phenomena. As Dahl remarks: 'The concepts of old and new information are used to explain such phenomena in language as intonation, stress and word order and the use of anaphoric devices' (1976: 37).

A scholar who has preserved the narrow interpretation of the status of given information and, indeed, attempted to redefine the term to force a narrow interpretation is Wallace Chafe, in a series of publications (1970, 1972, 1974, 1976). He writes:

The terminology has been and continues to be misleading to linguists and psychologists who use it. Calling something 'old information' suggests it is 'what the listener is expected to know already'. (1976: 30)

Chafe insists that given status should be restricted to 'that knowledge which the speaker assumes to be in the consciousness of the addressee at the time of the utterance' (1976: 30). He develops a series of images, such as 'in the forefront of the mind' (1970: 211) and 'spotlighted in the hearer's attention', in order to try to force home the *here and now* saliency which he attributes to given status at the moment of utterance. He stresses that givenness is a transient status: 'One indisputable property of consciousness is that its capacity is very limited. As new ideas come into it, old ones leave. The speaker's treatment of an item as given, therefore, should cease when he judges that item to have left his addressee's consciousness' (1976: 32). In Chafe's analysis, as in Halliday's, it is perfectly possible for a speaker to say *I saw your father yesterday*, where 'your father' is treated as new, if the speaker judges that the addressee's father is not in the addressee's consciousness at the time of the utterance.

A very different view of givenness is put forward by Herb Clark (in, for example, Clark & Clark, 1977). Clark takes the general view of 'given' status characterised by Chafe as 'what the listener is expected to know already'. Clark & Clark characterise information status thus:

given information should be identifiable and new information unknown . . . listeners should be confident that the given information conveys information they can identify uniquely. They understand that it is information the speaker believes they both agree on and that the speaker is asserting his beliefs about. (1977: 92)

This looks very like the characterisation of *presupposition* we provided in 2.1 (cf. also discussion in 5.2.2).

Perhaps the most expansive development of the Clarkian view of given status is to be found in Sanford & Garrod (1981). They propose a processing model which invokes the notion of *scenario* (discussed in 7.6). The scenario may be characterised here as a particular stereotypical configuration of past experience. Thus a courtroom scenario would be stored in memory with attendant typical characters and procedures. Sanford & Garrod suggest (1981: 114): 'The scenario enables referring to individuals to be made in the first instance by a definite noun-phrase, because they are already *given* in the representation.' Givenness is acquired then, not because this status is attributed to information by the speaker / writer, nor because it is salient in the context or previously mentioned in the co-text, but because of its status within the scenario which the language evokes. Thus, if a courtroom scenario is established, a 'lawyer' would form part of this scenario and hence could be treated as 'given'.

It is clear that the concept under discussion here is very different from that discussed by Halliday or Chafe. What is under discussion here has to do with the structuring of background knowledge (cf. 7.6). In a discussion of the sequences presented in (16b and c),

- b. 1. Mary got some beer out of the car.
2. The beer was warm.
- c. 1. Mary got some picnic supplies out of the car.
2. The beer was warm.

Chafe comments thus on Haviland & Clark's (1974) experiment, which showed that subjects took longer to process *c2* than *b2*:

In the first case, where the beer was mentioned in the context sentence, it is indeed given in the target sentence and we would expect that subjects, if they had been asked to read the target sentence out loud, would have pronounced it with low pitch . . . In the second case, where the beer was not mentioned in the context sentence, it is not given in the target sentence and subjects, if they had read it, would have been expected to give it high pitch . . . the basis for the definiteness of *the beer* in the two target sentences is not the same.

(1976: 41)

Chafe goes on to suggest that what Haviland & Clark are concerned with is not *givenness* (as the term is used by Chafe), but *definiteness* (1976: 42). He points out that givenness may coincide with

definiteness and often does, but that it is perfectly possible to find combinations of definiteness and newness (as in *c2* above) or as in *I saw the milkman yesterday for the first time for ages* said with no previous mention of 'the milkman' and without one in sight.

There is a great deal of room for confusion between the use of *given* as restricted to the status of information within a tone group (or clause) or extended to include whatever knowledge speakers and hearers share. In the next section we shall explore an attempt by Prince (1981) to develop a new taxonomy of information status, introducing new, and as yet unsullied, terminology.

5.3.2 A taxonomy of information status

If we wish to take account of the different views of information status expressed on the one hand by linguists like Halliday and Chafe, and on the other by psycholinguists like Clark & Clark and Sanford & Garrod, we need a richer taxonomy than the simple 'given / new' distinction that we have been working with so far.

Prince (1981) provides the basis for an extended taxonomy. She suggests that we should view a text as 'a set of instructions on how to construct a particular *DISCOURSE MODEL*. The model will contain *DISCOURSE ENTITIES*, *ATTRIBUTES* and *LINKS* between entities' (1981: 235). How is such a model to be constructed?

Prince suggests that the speaker may introduce new entities into the discourse. New entities are of two types. *Brand new* entities are assumed not to be in any way known to the speaker and will typically be introduced into the discourse by an indefinite expression like *a man I know* or *a bus in Princes Street*. The second type of new entity, *unused* entity, is assumed by the speaker to be known to the hearer, in his background knowledge, but not in his consciousness at the time of utterance. Chafe's example *I saw your father yesterday* (1976: 30) would fit into this category, as would expressions like *Chomsky* or *Jackendoff* addressed to a student of linguistics who, the speaker believes, is currently thinking about instrumental phonetics, rather than syntax, for example.

Prince calls her second class of entities *inferrables*. These are entities which the speaker assumes the hearer can infer from a discourse entity which has already been introduced. Thus 'the

driver' would be inferrable from the interpretation of the expression *the car*, as long as you have the background knowledge that 'cars have drivers'. There is no difficulty in interpreting the expression *the driver* in (26):

- (26) there was a car approaching the junction + but the driver didn't stop at the give way sign

In Prince's terms, it is an inferrable relationship which enables us to interpret the second sentence in (16c-f) as relating to the first sentence in each case. The class of inferrable entities would include, presumably, those scenario-based entities which Sanford & Garrod (1981: 114) classify as 'given' (e.g. courtroom scenario = lawyer).

The third set of discourse entities is the *evoked* class. There are two types. The first, *situationally evoked*, is salient in the discourse context (for instance 'I' and 'you'). The second is *textually evoked* and is an entity which has already been introduced into the discourse which is now being referred to for the second or subsequent time; (16a and b) exemplify this. 'Evoked' entities are what Halliday and Chafe expect to find speakers treating as 'given'.

We need to introduce a further distinction into Prince's category of *textually evoked entity*. In 5.2.1, we drew a distinction between *current* and *displaced* entities, both of which have already been introduced into the discourse, but where the *current evoked* entity is the one which was introduced as 'new' immediately before the current *new* entity was introduced. Displaced entities were introduced prior to that. This distinction, as we observed in the distribution of realisation forms in data, is justified by the difference in the typical realisation of the two categories.

The combination of Prince's system with our current / displaced distinction yields the following taxonomy of information status:

(27)	<i>New</i>	<i>Inferrable</i>	<i>Evoked</i>
	brand new		situational
	unused		textual - current
			- displaced

In section 5.3.3, we shall first try to apply this taxonomy within the restricted data we have already discussed, and then consider its wider applications.

5.3.3 The information status taxonomy applied to data

Within the restricted diagram-drawing data which we have already referred to in previous sections, it is possible to discern the different categories of entity identified by Prince, as they are introduced, and to observe the form of expression used to refer to them. A synopsis of the outcome of such an analysis is presented as (28). (Detailed discussion of different aspects of this analysis is to be found in Yule, 1981 and Brown, 1983.)

- (28) Forms produced to refer to:
1. NEW ENTITIES
 - a. *brand new*
 - (i) draw *a black triangle*
 - (ii) draw *a straight line*
 - (iii) write *OUT in black*
 - (iv) there's *a circle* in the middle
no examples in this restricted data
 - b. *unused*
 2. INFERRABLE ENTITIES
 - (i) it's right through *the middle* (circle)
 - (ii) you start at *the edge* (triangle)
 - (iii) with *the right-angle* (triangle)
 - (iv) *the corner* (triangle)
 3. EVOKED ENTITIES
 - a. *situational*
 - (i) in the middle of *the page*
 - (ii) *you've* got a triangle
 - b. *textual - current*
 - (i) to the left of the red line
about half a centimeter above
it
 - (ii) there's a black circle . . .
above *it* there's
 - (iii) draw a line in the middle
and above *it* write ON
 - (iv) it's a right angle triangle
. . . the bottom line of *the triangle*
 - (v) A. it's in red
B. \emptyset in red +
 - c. *textual - displaced*
 - (i) draw a black triangle . . .
underneath *the triangle*
 - (ii) to the left of *the red line*
 - (iii) *the black one*
 - (iv) at the base of *the red one*

From a syntactic point of view there are a number of forms of expression available to a speaker from which he can choose when he refers to an entity. We can summarise these informally thus:

- (29)
- | | |
|-------------|-------------------------|
| <i>a</i> | (+ properties) <i>X</i> |
| <i>the</i> | (+ properties) <i>X</i> |
| <i>it</i> | |
| \emptyset | (ellipsis) |

How are these forms distributed in our data? As you will see, *brand new* entities are regularly introduced by *a (+ properties) X*, where the specification of properties is not always present. The words *IN* and *OUT* in this data are referred to by naming the word - we shall not discuss the issue of names here. *Inferrable* entities are regularly introduced by definite expressions. In this data, inferrable entities include the known properties of circles (*middle, edge, radius, bottom*), of triangles (*apex, angle, side, and, for some individuals, right-angles*), of pages (*middle, corner, side, top*), or lines (*end, edge, top*), etc. *Evoked situational* forms are mostly used to mention the page which the hearer is drawing on, the hearer (*you*), and to the red and black pens which the hearer is drawing with.

Most of the expressions used to mention *current textual* entities are 'lexically attenuated', to use Chafe's term, either pronominal or elided, though there are some definite referring expressions. (We ignore here the problem of treating a null anaphor (or ellipsis) as a 'referring expression'. We assume that a structural gap is an instruction to the hearer / reader to fill it.) These entities, which the speaker knows have to be in the forefront of his hearer's consciousness, are referred to with the minimal referring forms. *Displaced textual* entities, on the other hand, are never, in this data, referred to pronominally or elided, but always referred to by a definite referring expression, often accompanied by an identifying property.

We may further note that whereas the expressions relating to *new* and *inferrable* entities are regularly realised with phonological prominence, 'evoked expressions' are not realised with phonological prominence except in a few cases of contrast. The interaction of distribution of syntactic form together with the distribution of phonological prominence significantly contribute to identifying the information status in terms of the taxonomy we have been discussing, as (30) shows.

(30)	Brand new	Inferr- able	Situa- tion	Current	Dis- placed
<i>a</i> + properties	77				
<i>a</i> - properties	21				
<i>the</i> + properties		29		3	44
<i>the</i> - properties	2	71	58	21	56
pronoun			9	65	
ellipsis			33	11	
prominence +	87	79	2		4
prominence -	13	21	98	100	96

(In each case the figures represent the percentage of expressions under each head which have this particular form. Scores for phonological prominence show percentage of actually occurring realisations.)

In 5.2.1 we discussed the striking difference in the characteristic syntactic form of expressions relating to current and displaced entities (except for the overlap of simple definite NPs). This difference appears to confirm the correctness of Chafe's view that givenness has a very transitory status. In this specifically transactional data, where the speakers introduce a number of similar entities, a displaced entity may be particularly liable to require specification of its properties in order to distinguish it from other, potentially competing, displaced entities. In spite of the increased specificity of the lexical identification, expressions used for displaced entities in this data are normally uttered without phonological prominence (not associated with an intonational peak), so are analysable, in Halliday's terms, as 'given'.

It should be pointed out that the interactions from which this data is drawn are, typically, very short, lasting no more than 3-4 minutes, and that the speaker knows, in each case, that the hearer has in front of him a physical representation of what has previously been mentioned, since it is the hearer's business to draw that representation. It might be claimed that everything which the hearer has drawn can safely be treated by the speaker as given. It is relevant to observe that it is only expressions of entities and properties of entities which have previously been *mentioned* in the discourse, which are phonologically non-prominent in this data. Those entities and properties which are classified, following Prince, as 'inferrables', are referred to with expressions which are phonologically prominent (on intonation peaks) hence, 'treated as new'.

This is in spite of the fact that the speaker is referring to 'ends' of lines, 'corners' of squares, 'angles' of triangles which the speaker knows are not only physically present in the context but have just been drawn there by the hearer. If the speaker has no reason to believe that the hearer is paying attention to these particular entities or properties of entities, he mentions them with phonological prominence. Once again Chafe's metaphors, which insist that the speaker must suppose that what he is talking about is *saliently* present in the hearer's consciousness, if he wishes to treat this information as given, seem particularly apposite.

Prince's taxonomy performs the useful function of distinguishing between what has been treated as 'given' in the linguistics literature ('situationally and textually evoked' mention, in her terms), as opposed to what has been treated as 'given' in the psycholinguistics literature ('situationally and textually evoked' mention *plus* the class of 'inferrables').

There remains, however, a real problem with trying to establish a taxonomy of information structure which can be appealed to in analysis independently of the forms which the speaker produces. Consider again some of the expressions which appear in (22), cited here as (31):

- (31) a. 'Slice the onion finely, brown in the butter and then place in *a small dish*
 b. . . . add to the fat in *the pan*
 c. . . . and boil for 20 minutes with *the lid* on the pan

None of the expressions italicised in (31) mentions an entity which has previously been explicitly introduced in the discourse. In each case this is the first occasion of mention. There is a sense, then, in which the items might all be held to be 'brand new'. On the other hand, it might be argued, the writer may reasonably suppose that this recipe is being read in the context of a kitchen, and he obviously believes that the reader will have in his kitchen the items 'a small dish', 'a pan' and 'a lid'. Is there not a sense in which all of these items are 'inferrables', introduced naturally from the 'kitchen scenario'? The reason why we feel that both of these arguments are misplaced is surely because these objects are not being *treated in the same way* by the writer. How do we know this? Because the forms of expression which he uses to refer to them by are different. The problem is that the only evidence we have of the information

status the writer attributes to different entities is the form of the expression which he produces.

5.4 Conclusion

We have tried, in this chapter, to give some indication of the many approaches to the question of 'information structure'. We have examined a range of forms, intonational and syntactic, which are held to be associated with 'new' or with 'given' information status. We have indicated that intonational prominence is not always associated with the use of an indefinite referring expression, and that lack of intonational prominence is not always associated with the use of a definite referring expression. We could show that indefinite referring expressions need not only be used on first mention (as in *draw a triangle + a triangle in red*) and that some first mentions take the form of definite referring expressions (as in 31b and c). We have pointed out that intonation prominence can be associated with other features of discourse than new information (for instance with contrast and emphasis) and that lack of intonational prominence is sometimes associated with first mention (discussed in 5.1.7). We have also pointed out that sentence form does not seem to determine given / new status, though it may indicate presuppositions on the part of the speaker. If we have to rely on linguistic forms alone to determine information status, it seems that the relevant status will not always be clearly marked and, indeed, if syntactic and intonational forms are both regarded as criterial for 'givenness', that these forms may supply contradictory information to the hearer.

We have compared the use of the term 'given' in the linguistics literature with its use in the psycholinguistics literature and attempted, by making appeal to the taxonomy put forward by Prince, to make clear how the term has been extended in the psycholinguistics literature. We considered whether, in principle, it is possible to establish a taxonomy of information status independently of the forms of expressions used by speakers, and we concluded that it is not possible. It seems that our only safe access to information status is provided by the form of the expressions used by the speaker / writer.

The discourse analyst might very well, at this point, decide to give up the concept of information structure since it is so hard to

pin down formally and impossible to pin down independently of formal expressions. We believe that the discourse analyst should avoid this counsel of despair. It is certainly the case, as Halliday has always insisted, that information status is determined, not by the structure of discourse but by the speaker. It is also certainly the case that there are no 'rules' for the specification of 'new' or 'given' status by the speaker. There are, however, regularities. In the diagram-drawing data we have described here, we can observe regularities which permit us to make statements like 'speakers usually introduce new entities with indefinite referring expressions and with intonational prominence' or 'speakers usually refer to current given entities with attenuated syntactic and phonological forms'. We must suppose that it is the exploitation of these regularities *in contexts of discourse* which allows us to assess the information status attributed to an entity by speakers and writers. As discourse analysts, when we examine data which is less rigidly controlled than that which we have analysed in detail in this chapter, we may lean on the regularities observed in the tightly controlled material to formulate an initial analysis. This methodology, which involves extrapolating from controlled data to uncontrolled material, provides a more secure basis for an account of information structure than can be produced by generalisations based only on short constructed sentences.